

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/140994/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Cavasso, Luca and Taboada, Maite 2021. A corpus analysis of online news comments using the Appraisal framework. *Journal of Corpora and Discourse Studies* 4 , pp. 1-38. 10.18573/jcads.61 file

Publishers page: <http://doi.org/10.18573/jcads.61>
<<http://doi.org/10.18573/jcads.61>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



LUCA CAVASSO
SIMON FRASER UNIVERSITY

MAITE TABOADA
SIMON FRASER UNIVERSITY

CITATION

Cavasso, L. & Taboada, M. (2021). A corpus analysis of online news comments using the Appraisal framework. *Journal of Corpora and Discourse Studies*, 4:1-38

KEYWORDS

Appraisal framework, evaluative language, online news comments, social media language

DOI

10.18573/jcads.61

ISSUE DOI

<https://doi.org/10.18573/jcads.v4>

ABSTRACT

We present detailed analyses of the distribution of Appraisal categories (Martin & White, 2005) in a corpus of online news comments. The corpus consists of just over one thousand comments posted in response to a variety of opinion pieces on the website of the Canadian English-language newspaper *The Globe and Mail*. We annotated all the comments with labels corresponding to different categories of the Appraisal framework. Analyses of the annotations show that comments are overwhelmingly negative and that they favour two of the subtypes of Attitude, Judgement and Appreciation. The paper contributes a methodology for annotating Appraisal, examining the interaction of Appraisal with negation, the constructive nature of comments, and the level of toxicity found in them. The results show that highly opinionated language is expressed as opinion (Judgement and Appreciation) rather than as an emotional reaction (Affect). This finding, together with the interplay of evaluative language with constructiveness and toxicity in the comments, can be applied to the automatic moderation of online comments.

CONTACT

Luca Cavasso and Maite Taboada, Department of Linguistics, Simon Fraser University, 8888 University Dr., Burnaby, BC, V5A 1S6, Canada

lcavasso@sfu.ca, mtaboada@sfu.ca

ORCID

0000-0003-3665-7920, 0000-0002-6750-8891

LICENSE

© The authors. Available under the terms of the CC-BY 4.0 license

A corpus analysis of online news comments using the Appraisal framework

Luca Cavasso
Simon Fraser University

Maite Taboada
Simon Fraser University

1. Introduction: Evaluative language in online news comments

Comment is a characteristic of contemporary life: it can inform, improve, and shape people for the better, and it can alienate, manipulate, and shape people for the worse. (Reagle, 2015, p. 183–184)

Evaluative content is one of the most prominent features of online and social media language. Whether positive ‘likes’ on Facebook or negative comments on a blog post, much of what we do online involves expressing emotions and evaluating ideas, people, and objects. Although there are many linguistic analyses on different aspects of online communication (Biber & Egbert, 2016; Clarke & Grieve, 2019; Dancygier & Vandelanotte, 2017; Hardaker, 2015, to mention just a few), little research has tackled the fine nuances of meaning in online language that can be explored through the Appraisal framework (Martin & White, 2005). We address this gap, examining evaluative language in online news comments, in particular comments posted in response to opinion articles. Appraisal provides a detailed characterization of the types of evaluation, beyond positive or negative, that can be found online. We explore a relatively small corpus through this qualitative, discourse-based framework. We show the potential that Appraisal exhibits for discourse-based corpus approaches and how important insights can be drawn even from small corpora.

Commenting is one of the fundamental features of the social media revolution. While the first wave of the internet was a one-to-many communication medium, the second wave, Web 2.0, emphasized social interaction (O’Reilly, 2005). This latest incarnation of the internet is still one-to-many (e.g., news organizations or bloggers posting content), but its distinguishing characteristic is that it is also many-to-one and many-to-many. Online participation involves posting one’s opinions for wide distribution, but also reacting to somebody else’s opinions and ideas. A large proportion of internet content in 2021 is user-generated and posted on large social media sites such as Facebook, Twitter, YouTube, Instagram, or TikTok.

Online comments arose out of this social media revolution, allowing readers to respond to online content. WordPress and similar blog hosting sites were the first to allow comments on blog posts. From there, they migrated to online news sites and many other forms of one-to-many content distribution. Research in journalism and public engagement has by now expanded to account for the role of comments, how news organizations view their usefulness (Manosevitch & Tenenboim, 2017; Reich, 2011; Stroud, Jomini,

Murray & Kim, 2020), and how they organize their moderation (Gillespie, 2018; Llansó, 2020; Loosen *et al.*, 2018; Roberts, 2019; Risch & Krestel, 2018; Risch, Repke, Kohlmeyer & Krestel, 2021; Wolfgang, 2018). Comments have been particularly controversial on news sites. Some news publishers have promoted and protected comments on their sites, arguing that they allow communication between readers, journalists, and the organization in general (Barber, 2017; Diakopoulos, 2015; Meyer & Carey, 2014; Meyer & Speakman, 2016). Other news sites, however, have found comments difficult to monitor and manage. Some of them have taken the drastic measure of closing comments altogether, among them *The Toronto Star*, *The Atlantic*, *Popular Science*, and National Public Radio in the United States, or closing them selectively (Bilton, 2018). Of particular relevance is the decision by the Canadian Broadcasting Corporation to not allow comments on stories relating to Indigenous issues, that is, relating to First Nations, Inuit, and Métis peoples of Canada, because those stories “draw a disproportionate number of comments that cross the line and violate our guidelines” (MacGuire, 2015).

The British newspaper *The Guardian* undertook an extensive analysis of the more than 70 million comments left on its site over a 10-year period (Gardiner, Mansfield, Anderson, Holder, Louter & Ulmanu, 2016). They had the unique ability to study comments that had been submitted, but which were never published on their site because moderators had identified them as abusive and had blocked them. Among their findings were that some sections and topics drew more abusive comments than others. World News, Opinion, and Environment had more abusive comments and so did, surprisingly, Fashion. But perhaps the most salient finding is that, of the top 10 most abused journalists (i.e., those that received the most abusive comments related to stories that they wrote), eight were women and two were Black men. There seems to be a bias against women and minorities, resulting in those populations receiving a disproportionate amount of abuse online (Hardaker & McGlashan, 2016; Jane, 2016; Nakamura, 2015; Sobieraj, 2020; Veletsianos, Houlden, Hodson & Gosse, 2018; Williams, Burnap, Javed, Liu & Ozalp, 2020).

The realization that abuse is common in some online communities has led to several strands of research describing, identifying, and automatically classifying comments based on their abusive or toxic nature (Benesch, 2012; Theocharis, Barberá, Fazekas & Popa, 2020; Warner & Hirschberg, 2012; Wulczyn, Thain & Dixon, 2017). Efforts in this area aim at automatically identifying abusive, toxic, and hate speech (Davidson, Warmesley, Macy & Weber, 2017; Nobata, Tetreault, Thomas, Mehdad & Chang, 2016; Waseem & Hovy, 2016), and at using counterspeech and different ways to present comments, in an effort to change the behaviour of abusive individuals (Kolhatkar, Thain, Sorensen, Dixon & Taboada, to appear; Risch, Repke, Kohlmeyer & Krestel, 2021; Wright, Ruths, Dillon, Saleem & Benesch, 2017).

Comments are interesting because of their use of emotionally charged and potentially abusive language. They are also a source of insight for news organizations and for news

readers alike, and thus an interesting register to analyze. In this article, we classify different types of evaluative language in news comments. Our goal is to categorize the relative distribution of Attitude types within Appraisal, establish how frequent negative evaluation is, and to map Appraisal annotations to independent annotations of the constructiveness and toxicity of the same comments. To our knowledge, this is the first study of online news comments applying the Appraisal framework. Our analyses reveal a spectrum, from constructive and insightful comments to trollish and abusive. Understand the nature of such comments and their linguistic characteristics is of interest in itself to linguists and corpus linguists. It can also be an important tool in the automatic and semi-automatic moderation of comments. We know that methods relying exclusively on keywords are not able to identify the worst abuse online (Benamara, Taboada & Mathieu, 2017); understanding the nuances of meaning expressed online would contribute to moderation platforms.

Before we introduce the data and the analysis, in the next section we discuss previous work on comments and describe how they have been studied. The rest of the paper is organized around a description of the theoretical framework, in Section 3, and of the data in Section 4. The main body of the paper explains our annotation of the data following the Appraisal framework (see Section 5) and the results of such annotation, in Section 6.

2. Online news comments: genre and characteristics

Previous studies of comments include general work on online content of various types, such as short reviews, responses in blog posts, or comments on news stories. A large body of research has examined the review genre in general, and reviews of book, films, and consumer products (de Jong & Burgers, 2013; Skalicky, 2013; Taboada, 2011; Vásquez, 2014). We will not focus on reviews, however, as the genre is quite different from the short comments that are typically found after news stories.

Reagle's (2015) extensive study of comments includes not only news comments, but the many forms that online reviewing and criticism may take, from the restaurant and film reviews already mentioned to comments on YouTube videos. Reagle's monograph provides an excellent historical overview of the commenting genre online, which he roots in the reviews provided by Michelin guides, albeit the study is sociological, and does not centre on the language of comments.

Comments have been studied in the computational linguistics literature, most commonly with the goal of classifying them as constructive (engaging, respectful, informative) on the one hand, or abusive on the other hand. Classification of abusive language is of paramount importance in automatic moderation systems, which aim to filter out posts that may be abusive, toxic, or constitute hate speech (Davidson, Warmesley, Macy & Weber, 2017; Kwok & Wang, 2013; Mishra, Del Tredici, Yannakoudakis & Shutova,

2019; Nobata, Tetreault, Thomas, Mehdad & Chang, 2016; Waseem & Hovy, 2016; Wulczyn, Thain & Dixon, 2017).

From a corpus linguistics point of view, studies of comments have focused on social media posts on platforms like Facebook, Twitter, or Reddit. Some of this work examines affiliation, identify, or conversational structure (Farina, 2018; Kiesling, Pavalanathan, Fitzpatrick, Han & Eisenstein, 2018; Theocharis, Barberá, Fazekas & Popa, 2020; Zappavigna, 2011). Of particular note is Zappavigna's (2012) analysis of Appraisal in tweets, a social media platform which she describes as being characterized by interpersonal meaning. Indeed, she finds that Twitter users often report on their own affectual state (Affect), with *I love you* as one of the most frequent collocations in the corpus. We shall see that, in our analyses, Affect was conspicuously absent. There are also excellent analyses of traditional media using Appraisal (Bednarek, 2006; 2010), but little research has been carried out using Appraisal to analyze the language of online news comments.

In the context of abusive language, it is worth mentioning the work of Hardaker (2015), who studied responses to trolling in community newsgroups. This seminal study categorized the ways in which communities organize themselves to protect against trolls and undesirable behaviour. Hardaker (2016) also examined an extreme form of abuse (rape threats on Twitter against a prominent feminist), meticulously documenting how online misogynist communities are built. Her analysis of the most frequent words and collocations in the corpus reveals the world view of those who post threats, but also the strategies used by supporters (e.g., the use of the phrase *real men*, as in *Real men don't rape*). Interestingly, Hardaker posits that users who posted offensive language, but not actual threats or illegal behaviour may easily escalate their behaviour in a context where it may become normalized. This is why analyses of such language, which may help in identifying abuse and threats automatically, are of paramount importance.

Some other corpus studies pursue a multi-dimensional lens, following in the tradition of Biber to study register variation (Biber, 1995), including our own analyses of online comments vs. traditional registers (Ehret & Taboada, 2020) and vs. other online registers (Ehret & Taboada, to appear). In these studies, lexicogrammatical characteristics are analyzed in a multi-dimensional space and mapped to dimensions of register variation. While many of these studies, such as Clarke & Grieve (2019)'s examination of Donald Trump's Twitter account, use corpus-linguistic methods to investigate comments on micro-blogging sites, their focus is not always on evaluative language. An exception is Berber Sardinha (2018), who discovered two different types of stance (evidentiality and affect) in a study of several online registers.

In terms of comments as a genre, most online genres have an origin in genres that existed well before the internet, such as the origin of email in the 20th century office memo (1992), or online blogs as an evolution from personal diaries (Giltrow & Stein, 2009; Herring, Scheidt, Bonus & Wright, 2004). Comments do seem to be an online genre exclus-

ively; their evaluative nature, however, can probably be traced back to a number of older genres, such as letters to the editor, reviews by professional writers in newspapers, or fan mail.

The news comments in our study, and news comments in general, escape characterization as a genre from a structural point of view, in the sense of genre from the literature that defines it as a goal-oriented activity which develops in stages (Martin, 1984; Martin & Rose, 2008). Since comments are so varied, it is difficult to identify stages, i.e., obligatory and optional parts that need to be in a comment to be recognized as such. The other aspect of a genre definition is the purpose of the genre. In this case, news comments can be defined as fulfilling a need to react, elaborate on, or contribute to ideas present in the article in question. They are dialogic in nature, since they are always a reaction to the article, other comments, or some combination of the two. The comments are, in fact, a form of polylogue (Marcoccia, 2004), a form of online communication with multiple levels of dialogue and different levels of participation.

Example (1) presents an example of this dialogic nature and gives a sense of the length and style of the comments. In this example, the dialogue is with the article, which discusses the issue of violence against Indigenous women in Canada (Turpel-Lafond, 2014). The author of the comment directly mentions the article (*this story*) and engages with it by articulating what they view as a solution to the problem.

- (1) This story gives broader context to the earlier reports of the abuse of band finances by the native leadership. The hardship on the reserves does have to be addressed but the people who have to lead on this are the reserve leaders themselves. Trying to do anything from the outside, as was the case recently with education, will be another waste of time and scarce money, resulting more bruised feelings. We have to see the problem for what it is: reserve based and nurtured.

In terms of register, that is, in terms of lexicogrammatical characteristics of the language of news comments, the defining characteristic of online news comments is the presence of evaluative language. That is precisely what we set out to study in this paper. In the sections that follow, we introduce the data, the analysis methodology, and the results of this study.

3. The Appraisal framework

Appraisal refers to a framework for understanding, classifying, and describing the linguistic resources deployed to express evaluation. It belongs in the systemic functional tradition of Halliday (Halliday, 1985; Halliday & Matthiessen, 2014), and is located in the discourse semantic stratum, the part of language concerned with meaning beyond the clause. Appraisal is both a linguistic system of meanings for evaluation (what language users deploy to make meaning) and a description of resources for evaluation (what linguists apply to analyze text) (Martin, 2017). We will use ‘Appraisal’ and ‘Appraisal frame-

work' most often in the second sense, but always bearing in mind that what we are trying to capture is how language users make (evaluative) meaning. The main tenets of the approach are discussed in Martin & White (2005), although many other publications exist, about different types of texts and different languages (Achugar, 2008; Becker, 2009; Lam & Crosthwaite, 2018; Taboada, Carretero & Hinnell, 2014).

Appraisal is concerned with how we adopt a subjective presence in language, how we use language to evaluate others, the world around us, and to express our own feelings. Following the systemic functional approach, it characterizes the linguistic resources for evaluation as a system of choices, a set of categories that speakers/writers choose from as they express evaluation, as shown in Figure 1, with some example realizations. The three main classifications break evaluation into Attitude, Graduation, and Engagement. Attitude is sometimes referred to as ways of feeling (Martin & White, 2005), and it constitutes the most salient classification of evaluation into emotion, ethics, and aesthetics (Affect, Judgement, and Appreciation). Affect captures emotional responses, either on the part of speaker or somebody else (sad, cheerful, anxious). Judgement refers to how people appraise other people, in terms of their abilities or behaviour, and whether it accords with moral and legal norms (kind, powerful, corrupt). Finally, Appreciation is the more general evaluation of objects from an aesthetic point of view.

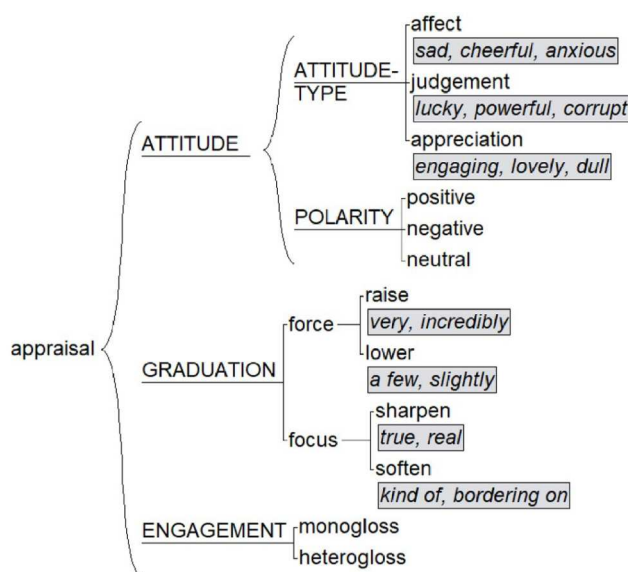


Figure 1: The Appraisal system

The description above, and the examples in the figure include mostly adjectives conveying the specific type of evaluation. Attitude, however, draws from all levels of language, from morphology (suffixes like *-let* or *-ish*) to discourse, including full sentences that encode evaluation. The examples in (2), all from our corpus, provide illustrations of each

type of Attitude, in brackets.¹ For Affect, the evaluation is jointly conveyed by the adverb *happily* and the verb *enjoying*. The second example shows a noun (*gibberish*) conveying Appreciation, and an entire sentence expressing Judgement, about the editor's competence.

- (2) a. I will bet he is [happily enjoying]_{Aff} every single day.
 b. This article is [gibberish]_{App} [How did it get past the editor?]_{Judg}

Attitude can, naturally express positive or negative evaluation. Thus, in the Attitude system, we include two choices: one about the type of Attitude and one about polarity. For instance, in Example (3), we see positive Affect in (3a) and negative Judgement for all the items in (3b). We found very few cases of neutral polarity and, upon reflection, most were assigned to either positive or negative, given the context. In (3c), the context was not sufficient; it is hardly an endorsement to say that a policy is not racist, and the rest of the comment evaluates other commenters rather than the legislation, so this expression was annotated as neutral.

- (3) a. Even as she [fondly reminisces]_{Aff, pos} here about smoking pot with her
 b. global [kings]_{Judg, neg} who [wish to break us down]_{Judg, neg} so they can [steal]_{Judg, neg} our resources
 c. There is [nothing racist]_{App, neu} about this piece of legislation, period.

The second system in Appraisal, Graduation, concerns how we amplify or downtone Attitude. We were interested in annotating Graduation, as social media has been described as having a tendency to upscale (Zappavigna, 2012, p. 67). Martin & White (2005) establish two types of graduation: Force and Focus. With Force, the emphasis is on placing the evaluation in some sort of scale, applying to words that are intrinsically gradable or can be made so. Focus tends to be used with non-gradable items, highlighting their prototypicality or fitness in a reference group. In the original description of the theory, Force can take the form of intensification (*very, slightly*) or quantification (*small, a few*), and each, in turn, can graduate up or down (*very* vs. *slightly*). Focus is divided into sharpen (*a true friend*) or soften (*an apology of sorts*). To simplify our annotation, we decided to assign values of 'raise' and 'lower' (or 'up' and 'down') for both Force and Focus. In (4), we show all four combinations, with the Attitude in brackets, and the specific word that conveys Graduation in italics.

1 All examples in the paper are from our corpus and are reproduced as they appeared, including typos, misspellings, and non-standard grammar. We, of course, do not endorse or condone the views expressed in the comments. We use brackets to indicate which parts had an Appraisal label. Abbreviations: Aff = Affect; Judg = Judgement; App = Appreciation; pos = positive; neg = negative.

- (4) a. Their goal is [*completely unbridled*]*Judg, neg, Force, raise* fossil fuel exploitation.
 b. you will have [*some credibility*]*Judg, pos, Force, lower*
 c. a [*truly loving*]*App, pos, Focus, raise* relationship
 d. [*kind of funny in a way*]*App, pos, Focus, lower*

Graduation always depends on an element that is already labelled for Attitude, that is, it does not occur in isolation, but instead only in the context in which something is being evaluated as Affect, Judgement, or Appreciation.

The last Appraisal system is Engagement, a set of resources for expressing the speaker's attitude to the evaluation itself and presenting it as open to negotiation or not. When there is no possible negotiation, the evaluation is monoglossic, that is, it is presented as non-negotiable. This is typical of statements without hedging or modulation. In heteroglossia, following Bakhtin (1981), positioning is open. Heteroglossic utterances involve a dialogic perspective, an acknowledgment of prior utterances, possible alternative viewpoints, and anticipated responses. A monoglossic utterance such as *The banks have been greedy* can be turned into utterances that recognize heteroglossic perspectives, such as *In my view, the banks have been greedy* or *There can be no denying the banks have been greedy*.² Engagement is a complex system of resources and choices, and we did not explore Engagement annotations in our study, but, naturally, the examples we show here sometimes contain instances of Engagement, in particular when it comes to negotiation (see Section 6.3).

The examples in this section all include instances of evaluative language where the evaluation is clearly attached to a specific word or phrase. Those are defined as *inscribed* Appraisal, that is, evaluation that is directly realized in the language through the use of attitudinal expressions. In many cases, however, evaluation is *invoked*, in which “an evaluative response is projected by reference to events or states which are conventionally prized” (Martin, 2000, p. 142). We consider both inscribed and invoked Appraisal. We annotated invoked Appraisal whenever context allowed, as in (5). In (5a), it is understood that young people going missing is bad, especially in the context of the article, which discussed crimes against Indigenous youth such as murder and abduction. In (5b), the commenter's description invokes the father's tenacity, reflecting positively on his character.

- (5) a. [60,461 youth were reported missing]*App, neg*
 b. My father could not read nor write, but he [worked very hard and managed to land a unionized job]*Judg, pos*

2 The examples in this paragraph are from Martin and White (2005, p. 100).

Since Appraisal was first proposed (Eggins & Slade, 1997; Horvath & Eggins, 1995; Martin, 2000; White, 1998, 2002, 2003) and formalized in Martin & White's (2005) book, a substantive body of work has developed, extended, questioned, and applied different aspects of the framework to an extensive range of texts. It is worth mentioning here that, although the expression 'Appraisal Theory' has been used in print, Martin has stated that he views it as a framework, rather than a theory: 'Systemic Functional Linguistics [...] is the theory. Appraisal is a description of resources for evaluation in English' (Martin, 2017, p. 22).

Appraisal lends itself well to the study of different types of evaluative text, including extensive work on political discourse and news stories (Coffin & O'Halloran, 2006; White, 2002, 2003, 2016), narratives (Macken-Horarik, 2003), casual conversation (Eggins & Slade, 1997), movie reviews (Taboada, Carretero & Hinnell, 2014), or wine reviews (Hommerberg & Don, 2015).

Of particular note is the work of Fuoli on company corporate responsibility reports (Fuoli, 2012) and CEO letters (Fuoli & Hommerberg, 2015), exploring trust and transparency in communications between companies and the public. An important contribution in this line of work is Fuoli (2018), which presents a detailed method to devise, carry out, and explore Appraisal annotations. One of the challenges in annotating and analyzing Appraisal is that the analysis can be perceived as subjective, because interpretations of evaluative content are very much context-dependent (Ben-Aaron, 2005; Hommerberg & Don, 2015; Macken-Horarik & Isaac, 2014; Thompson, 2014). Fuoli's stepwise method addresses these problems by ensuring that the annotation is transparent and reliable. We have, in the annotation for this paper, also developed clear guidelines, and have conducted two reliability studies (see Section 5), following principles developed in our previous work annotating online reviews (Taboada, Carretero & Hinnell, 2014). The guidelines are available with the public release of the larger SOCC corpus (Kolhatkar, Wu, Cavasso, Francis, Shukla & Taboada, 2020), including the raw and annotated versions (see next section).

4. Data: The SFU Opinion and Comments Corpus

As part of a large project on the nature of evaluative language, we are exploring the evaluative content of online news comments and, for that purpose, collected a large dataset of opinion articles and all their comments from the website of the Canadian English-language daily *The Globe and Mail*, a relatively high-brow, business-oriented newspaper and, arguably, the paper of record across Canada. The data includes all opinion articles posted for the five-year period between 2012 and 2016. In addition, we collected all the comments relating to those articles. This larger corpus is the SFU Opinion and Comments Corpus, SOCC (Kolhatkar, Wu, Cavasso, Francis, Shukla & Taboada, 2018), publicly available and described in Kolhatkar *et al.* (2020). SOCC was downloaded under the 'fair

dealing' provision of Canada's Copyright Act, which allows download of copyrighted material for research and study purposes. It comprises three key components: articles, comments, and threads. The articles are columns, op-eds, and newspaper editorials published between January 1, 2012 and December 31, 2016, a total of 10,339 articles. The comments corpus contains all comments posted in response to those articles, a total of 663,173 comments. The comments corpus simply contains comments in sequential order. The threads corpus organizes comments by the groupings in threads under which they were posted, preserving reply structure. Table 1 provides a summary of all the components of the corpus. More information on the corpus, and the data collection corpus, can be found in Kolhatkar *et al.* (2020).

	Item	Count
Articles corpus	Number of articles	10,339
	Number of words in articles	6,666,012
	Number of unique article authors	1,628
	Average number of comments per article	85
Comments corpus	Number of comments	663,173
	Number of words in comments	37,609,691
	Number of unique commenters	34,472
Annotated Appraisal corpus	Number of comments	1,043
	Number of words	64,792

Table 1: Statistics of the SFU Opinion and Comments Corpus

From this large corpus, we extracted a subset of comments to annotate. Given the intensive nature of Appraisal annotation, only 1,043 were annotated manually. These were selected by extracting the top 100 comments or so (preserving thread structure) from 10 articles covering topics such as Indigenous relations, the federal budget, relations with China, a proposed national daycare plan, or property taxes.³

The Appraisal annotation is one of several manually labelled subcorpora, which include negation and its scope, constructiveness, and toxicity. We performed several annotations because we are interested in the interplay of all these characteristics. In this paper, we describe the Appraisal analysis of this subset of the corpus, and how it interacts with negation, constructiveness, and toxicity. The larger SOCC, and the smaller subset analyzed in this paper are instances of a corpus in the corpus linguistics sense, that is, in the sense that they are collections of language occurring in context and, as such, are suitable of analysis with corpus-based discourse analysis methods such as the ones presented

3 The original 10 articles are available on our lab webpage: <https://www.sfu.ca/discourse-lab/research/globe.html>

here. Corpus-based (or corpus-assisted) discourse analysis typically relies on concordances, collocations, and keywords to explore different types of text (Baker, 2020; Flowerdew, 2013; Partington, Morley & Haarman, 2004), especially how social phenomena are enacted in discourse (Baker, 2014; Baker, Gabrielatos & McEnery, 2013) and how evaluative prosody can reveal the discourse properties of words and expressions (Partington, 2014). In this paper, we push that discourse analysis beyond the boundaries of collocations, by exploring evaluative expressions labelled with Appraisal categories. The next section provides a detailed account of the annotation process.

5. Analysis

We followed a carefully designed approach to annotation, starting with an extended process of developing and testing guidelines, which was carried out with the help of members of our research group. Although Appraisal as a framework is well defined, labelling individual cases becomes complex, because many decisions are somewhat subjective, such as how much context is necessary for interpretation or how much of the evaluative expression should be labelled. This is why automatically labelling Appraisal accurately is not currently feasible (Dotti, 2013; Read & Carroll, 2012; Taboada & Grieve, 2004).

The annotation followed two main guiding principles: minimality and contextuality. Accountability was also very important, which is why we engaged multiple annotators and performed reliability tests throughout.

The principle of minimality means that the item to be annotated (henceforth a *span*) should be as short as possible, while at the same time including all the words that convey Attitude. This leads to spans of varying length, from single words (6a) to constituents (6b)⁴ and entire sentences (6c). Note that (6c) has two separate spans, co-dependent on each other for complete interpretation. This is a complex example, which could be considered invoked evaluation. To make it inscribed, that is, an instance we annotated and included in the corpus, we had to include large spans.

- (6) a. in these [desperate] times, those with. . .
- b. [another soar loser] who lives in the cry baby world
- c. [Don't complain that other women are running for office and getting elected] - [try it yourself and then complain if you think it's unfair].

This process of deciding what is the unit of analysis is one of the most challenging aspects of linguistic annotation. This process has been referred to as *identifying markables* or *spans* (Taboada, Carretero & Hinnell, 2014), or as *unitizing* (Artstein & Poesio, 2008; Fuoli, 2018).

4 Reproduced verbatim (*soar* for *sore*).

The second general principle for annotation is context dependence. This involves using any information available to understand the meaning of the evaluative expression under consideration. Annotators read the article that the comment was posted in response to and were also encouraged to draw from their own experience, of the world and of online language, to decide on the length of the span and assign the most likely label. For instance, in (7), *a selfless angel of mercy* could have been either positive or negative. We know, however, that the rest of the text disparages the Monsanto company. We also rely on the linguistic context and interpret the *sooo* at the beginning as a marker of sarcasm.

(7) Sooo, Monsanto is now a [selfless angel of mercy]?

The full annotation guidelines, with numerous examples, are available from the corpus description page (see footnote 7). In the rest of this section, we outline the general principles for classifying different types of Attitude, and for how to label Graduation.

5.1 Annotating Attitude

The theoretical distinction among the three types of Attitude (Affect, Judgement, and Appreciation) is quite straightforward. Affect refers to the expression of the speaker's feelings and emotions, or the description of somebody else's feelings. Judgement is used to evaluate people, especially their behaviour, morals, ethical characteristics, or capabilities. Finally, Appreciation occurs when we assess objects from an aesthetic point of view. In practice, however, there are many cases where the categories overlap, and a great deal of cultural and contextual knowledge is required to discern the nature of the evaluation.

It is particularly difficult to distinguish Judgement from Appreciation. Martin & White (2005) argue that this is because they are both in a sense derived from the more basic Affect. It is likely that we first developed a language for discussing emotions, and then reused that for other forms of evaluation. Judgement and Appreciation are, then, extensions, one dealing with ethics and the other one with aesthetics. The crucial distinction comes when describing organizations. They are things, abstract entities, but, at the same time, they are headed and administered by people, whose behaviour can be judged. In our guidelines, we suggest that a company, organization, or government may be appraised as if it is a thing (Appreciation) or a group of people (Judgement). As a general test, if a word implies agency or intent, it is probably an instance of Judgement. For instance, in (8), the commenter describes the Chinese Communist Party as *brutal* and *blood-thirsty* and, as both descriptions require some sort of agency, it is clear that the commenter is describing the members of the Party, and therefore this is an example of Judgement rather than Appreciation.

(8) The [brutal] Chinese Communist Party has [murdered] over fifty million of its own people since 1949, since 1999 it has been attempting the [blood-thirsty genocide] of the tens of millions of innocent Falun Gong who live in Mainland China.

On the other hand, Example (9) is an instance of Appreciation. *High unemployment* is an element of Canada's economy, not part of the character of its leadership or people.

(9) Canada has a [high unemployment] rate

In certain cases, spans seem to contain two types of Attitude. In our annotation, we tried to determine which one seemed primary. In Example (10), *hurt and domination* are primarily Appreciation, about the quality of a relationship, but there is also some affective content. Similarly, in (11), the phrase in brackets is negative Appreciation of the goal of exploitation, but, in the context, the sentence conveys negative Judgement of those exploiting fossil fuels.

(10) [Hurt and domination] has no place in a truly loving relationship.

(11) Their goal is [completely unbridled fossil fuel exploitation.]

Each instance of Attitude was also annotated with polarity, whether positive, negative, or neutral. Our instructions to annotators were to include as much context as necessary to determine the polarity of a particular span. In many cases, polarity was determined by the general tone of the whole comment, a case of semantic or discourse prosody, where the positive or negative connotations of the context affect individual words and phrases (Louw, 1993; Partington, 2014; Stewart, 2010).

Incorporating context was often necessary to detect sarcasm, as in (12), where the phrase *thank you* is clearly not genuine. As well, many comments feature pointed rhetorical questions such as (13), a comment on an article titled "Why Belgium is ground zero for jihadi terrorism" (Gagnon, 2016). The comment might be interpreted as asking genuine questions, except for the fact that *breeding ground* is in scare quotes, and engaging with the questions makes it clear that the commenter is trying to undermine the idea that Belgium provides a significant source of Islamic terrorism. Context is also necessary to determine whether political adjectives such as *liberal*, *conservative*, or *socialist* are intended to convey negative (or positive) Appraisal, such as in (14), where it is clearly negative, and especially evident by the rhetorical question at the end.

(12) This article was a big disappointment. Thank you Ms Henein. Now women know reading your emotion-based opinion piece is not an option.

(13) What is this bigger "breeding ground" that you speak of? Of all the terrorist acts committed in the last ten years, how many were perpetrated by Belgian Muslims?

(14) The NDP want kids in a unionized environment from birth to the end of university. And then ideally as voters they will support the NDP's socialist agenda. What could go wrong?

We found few cases where an annotation of 'neutral' was justified. Some of them involved negation of a negative statement, to diminish the negative meaning, while at the same time not stating a positive, as in (15) and (16). The concept of neutral evaluation

may sound like an oxymoron and, indeed, it may not be evaluation that is involved here, but rather Engagement. Since we were concentrating on annotating Attitude, we allowed annotators to apply a neutral label where a simple positive or negative assessment did not seem appropriate, as a case of ambiguous polarity.

(15) This is [not an embarrassment].

(16) But there's also [nothing wrong] with wanting to do things 'the new way' because we all did things in new ways at some point in our lives.

5.2 Annotating Graduation

Graduation is only annotated within a span that has Attitude, i.e., Graduation never occurs by itself. We also tried to restrict Graduation to the specific item that conveys it, rather than labelling the entire Attitude span as containing some Graduation somewhere. For instance, in Example (17),⁵ the span *completely ignored* is an example of negative Judgement (the journalists did not do their work), but the only word labelled with Graduation is *completely*, italicized in the example, because it is the one that primarily conveys the intensification.

(17) Meanwhile, our so-called journalists have [*completely ignored*] another officer involved shooting that occurred on August 11th.

Graduation is of two main types, Force and Focus (see Section 3). Force implies gradability, in scale or quantity, and we labelled it as “up” (18) or “down” (19).⁶ In the examples, the entire span is in brackets, and Graduation is italicized.

(18) The column is [alarmist *in the extreme*].

(19) The full arctic is [not *ALL* Canadian].

Focus applies to non-gradable items which are evaluated based on fit or prototypicality with respect to a class. For instance, in (20), the correctness is assessed as not open to discussion (and thus overlapping with Engagement). In (21), we see a mix of two types of Focus: *certainly* sharpens the expression, but *almost* dampens that assessment.⁷

(20) So, when the Chinese claim that the West is applying 'double-standards' they are, [*unquestionably*, correct].

(21) Elizabeth Warren, to take one example, [*almost certainly* would have produced a different outcome].

5 There are other spans with Attitude in the example, but they are not marked here because we use the example for illustrative purposes.

6 The negation in Example (19) can also have an Engagement reading, as a disclaimer. We did not annotate Engagement in our project (see Section 6.3).

7 As with Example (19), some of the expressions in these examples (*unquestionably*, *almost certainly*) are also expressions of Engagement, which we did not annotate.

5.3 Interannotator agreement

The annotation guidelines and the general framework for annotation were developed by the two authors. Then, to ensure that the guidelines were transparent, and to have a good assessment of how complex the annotation was, we hired a research assistant to perform the full annotation. She first worked with one of the authors, reading over the guidelines and performing multiple tests on a small number of comments. Once we felt she was ready to annotate, she annotated on her own, checking with us on a regular basis. The annotation process also involved two checks of interannotator agreement, at the beginning and at the end. Once the research assistant had annotated 50 comments, one of the authors annotated the same comments independently, and we checked agreement. A new set of 50 comments was compared in the same fashion, towards the end of the project. Finally, once all the annotations were completed, one of us curated the annotations, examining each one and ensuring it was accurate, and making any corrections when necessary. The annotation process took approximately three months, with the guidelines having been developed over a few months prior to that.

The agreement comparison was performed by calculating agreement based on labels (Attitude and Graduation), and the subcategories for those (polarity for Attitude; Force/Focus and up/down for Graduation). We also included length of span in our calculations. Full agreement consists of agreement on where the annotation begins and ends, and the label. When one annotator selected a slightly different portion of the example as compared to the other, we considered that as a disagreement. Agreement is calculated as a percentage. We did not employ more complex measures such as Cohen's kappa or Krippendorff's alpha, because most of our labels are binary and a percentage agreement suffices for such cases. Additionally, using any chance-corrected agreement measure on decisions such as the length of a span results in agreement that rapidly approaches zero as comment length increases. The results of both studies are shown in Table 2.

		Attitude	Graduation
First 50	Category agreement	81%	35%
	Polarity agreement	89%	46%
	Average	85%	41%
Last 50	Category agreement	81%	42%
	Polarity agreement	87%	48%
	Average	84%	45%

Table 2: Results of agreement study

Full details of the agreement study, including the first set of comparisons, can be found in our corpus description paper (Kolhatkar *et al.*, 2020). Here, we discuss only general areas

of disagreement. Many of the disagreements involved length of spans, that is, the process of identifying markables. An example is shown in (22), where one annotator began the span at *liberal* and the other limited it to *fear mongering and hate*. Since the comment later mocks *liberal 'logic,'* it was determined that the commenter likely considers the word *liberal* to be inherently negative.

- (22) So all the [liberal fear mongering and hate] directed at Harper for a deficit, now doesn't really matter when a surplus is on the books. Who needs a cat and a laser pointer when you have liberal 'logic' to entertain!

The other general source of disagreement was about labels, especially between Appreciation and Judgement, such as in (23). In this comment, the commenter criticizes the *response to radical Islam* as well as those taking it. We decided that in this case Judgement (that the European Union is implied to be not courageous or proactive enough in its response to terrorism) is more salient, since the EU is specifically named and the article is about the growth of terrorism in Europe.

- (23) The EU response to radical Islam:#JeSuisYourTownHere

As is clear from Table 2, the level of agreement for Graduation is quite low. While we use those annotations in our analyses in the next section, we rely mostly on the annotations for Attitude, which show moderate to high agreement.

6. Analysis of the annotations

The annotations were performed with the WebAnno annotation tool (de Castilho *et al.*, 2016), which not only provided an annotation interface, but also a way to curate and compare data from multiple annotators. The output of WebAnno was imported into comma-separated value files (CSVs) using Python and the Pandas package (McKinney, 2010). Finally, the statistical programming language R was used to run all the analyses (Mullen, 2016; R Core Team, 2018; Wickham, 2009). The scripts used for the analysis are available from the corpus download link (Kolhatkar *et al.*, 2018).

In this section, we provide first an overall summary of the frequency and distribution of Appraisal in the corpus, focusing on Attitude labels (Affect, Judgement or Appreciation) and polarity (positive, negative, neutral). We then move on to more detailed analyses of how labels pattern within comments, and the interaction of Appraisal with three other types of annotations that we performed separately: negation, constructiveness, and toxicity.

6.1 Overall frequency and distribution

The corpus comprises 1,043 comments, 3,973 sentences and 64,792 words. The number of spans of each label and polarity of Attitude are shown in Table 3. In terms of polarity, negative spans are overwhelmingly frequent, making up 4,867 or 73.5% of Attitude ex-

pressed in the comments. Positive spans make up almost the entire remainder (25.5%) of Attitude spans, with neutral Attitude only expressed in 1% of spans. As for the Attitude label, comments were about evenly split between Appreciation (54%) and Judgement (43%), somewhat favoring Appreciation. Meanwhile, Affect was quite rare, comprising only 3% of spans.

	Affect	Judgement	Appreciation	Total
Negative	175 (77%)	2,342 (83%)	2,350 (66%)	4,867 (73.5%)
Positive	46 (21%)	469 (17%)	1,173 (33%)	1,688 (25.5%)
Neutral	5 (2%)	10 (0.4%)	53 (2%)	68 (1%)
Total	226 (3%)	2,821 (43%)	3,576 (54%)	6,623 (100%)

Table 3: Instances of Attitude. Percentages (rounded off) for non-total rows are out of the column total. Other percentages are out of all spans.

The low levels of Affect are worth pointing out. One may think that online discourse involves references to emotions and emotional behaviour and that commenters typically express their opinion as a description of their emotions, using the first person (*I like this candidate; I am outraged at this situation*). This is not at all what we found in our data. It is rather the case that Affect is rarely used and opinion is instead conveyed through Judgement (*The candidate is accomplished*) or Appreciation (*The situation is outrageous*). This is a form of the ‘Russian doll’ phenomenon that Geoff Thompson pointed out (Thompson, 2014), whereby an expression of one type of Attitude functions as an indirect expression of another type. In this case, Judgement or Appreciation possibly being used as an indirect expression of Affect.

Due to this conflation of different types of Attitude, researchers have proposed a reorganization of the labels. For instance, Bednarek (2009) suggests that the evaluative space be divided into two main categories, Emotion and Opinion. Thus, Emotion includes the basic emotion categories (*happy, sad*), whereas Opinion focuses that evaluation on people and objects in terms of ethical or aesthetic norms (Judgement and Appreciation). The key innovation in Bednarek’s proposal is that many cases include a double coding of both Emotion and Opinion, as they may both convey affective content with the opinion. Benítez Castro & Hidalgo Tenorio (2019) explore this distinction further, refining the Emotion category (the original Affect in Appraisal) and grounding it in psychological principles. Our corpus results seem to support this reorganization, in particular with regard to the double coding of some instances of Opinion (Judgement and Appreciation in our analyses). In retrospect, we could have probably double-coded some of those as also conveying Emotion, in terms of the highly involved nature of the opinion.

In general, the trends for polarity hold within each label, and vice versa, but not to the same extent. Appreciation seems to lean positive: 33% of Appreciation spans were positive, as opposed to 17% of Judgement spans and 25.5% overall. Nevertheless, the vast ma-

jority of spans for any label is negative. This seems to contrast with studies that show that some genres, including online genres like movie reviews, tend to have more positive than negative words (Potts, 2011). The Hedonometer project (Dodds *et al.*, 2015) has shown a higher frequency of positive terms on Twitter. We believe that this higher frequency of negative Appraisal may be another characteristic of the genre of online news comments.

Virtually all comments contain some form of Attitude. The three (0.3%) comments that were not annotated as containing Attitude are provided below.

(24) !!!!!!!!!!!!!!!

(25) Ma lines is not a suburb of Brussels, it is the French name for an old Flemish city called Mechelen and is a good 30 km away from Brussels.

(26) Sorry. I meant a water pipeline from Canada to California.

In (24), the comment was judged to be too ambiguous to annotate, but was likely expressing either strong agreement with another commenter or with the article, or expressing surprise at another comment or the article and thus, in fact, was likely meant to convey some sort of Attitude, but we could not decide which. One could consider (24) as an instance of Graduation, as typography often fulfills that role (cf. Zappavigna, 2012), but we did not annotate any Graduation in the absence of Attitude.

The result is that 99.7% of comments in our corpus contain some sort of Attitude, which shows that, as a genre, their purpose is evaluative. This idea is further supported by the fact that commenters frequently expressed Attitude multiple times within one comment. For comments with at least one span of Attitude, each comment had a mean of 6.4 Attitude spans in it, and a median of 5.

Table 4 shows the number of spans of each type of Graduation. Graduation clearly trends towards upscaling by Force (quantitative or gradable intensification). Of all the Graduation spans, 91% were upwards Graduation and 85% of those used Force. Given the small amount of Focus and downwards Graduation, it is hard to make any useful observations about their distribution. Recall also that the inter-annotator agreement for Graduation is quite low. Although we curated the final set of annotations, we report these results as preliminary.

	Force	Focus	Total
Up	619 (80%)	79 (10%)	698 (91%)
Down	34 (4%)	39 (5%)	73 (9%)
Total	653 (85%)	118 (15%)	771 (100%)

Table 4: *Instances of Graduation*

Only 398 (38%) of our comments had at least one Graduation span. The majority of the Attitude spans occurred without Graduation. The rarity of Graduation was contrary to

our expectations. We expected online news comments to be highly opinionated, and to use Graduation to intensify and highlight those opinions. It seems rather that analytic forms of Graduation are eschewed, perhaps in favour of infused Graduation, where the Graduation is conveyed by a lexical choice rather than an intensifier, which we did not annotate. In other words, *very good*, an analytic form, may be less frequent than *amazing*, which contains infused Graduation. Since we did not annotate individual Attitude words such as *amazing* with respect to some scale of Graduation (*good* – *great* – *amazing*), we do not have the data necessary to explore this question.

6.2 Patterns within comments

Comments are overwhelmingly comprised of solely negative Attitude. We illustrate this with Figure 2, a density plot showing the percentage of spans per comment. A majority of comments had 0% positive spans. A total of 45% of comments contained only negative Attitude (measured in number of spans per comment), while 79% contained mostly (or entirely) negative Attitude. Eight percent were evenly split, 13% were mostly (or entirely) positive, and only 5% were purely positive. It seems that, in addition to being evaluative, a defining characteristic of this genre is that such evaluation is distinctly negative.

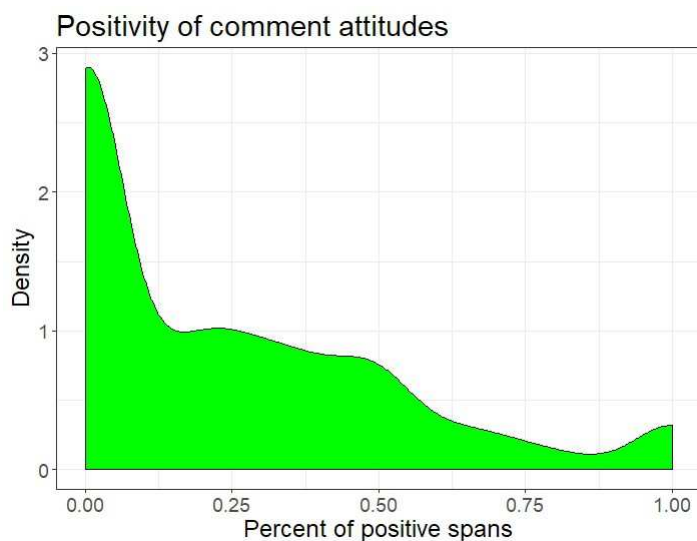


Figure 2: Density of positive spans per comment

The trend towards negative Attitude manifests more strongly per comment than in the corpus as a whole. Counting comments as positive if they consist of more than 50% positive spans and negative if they contain more than 50% negative spans, we find that 79% of the comments in the corpus were negative, as compared to 74% of the spans (cf. Table 3). Mostly positive comments were rare: 13% of the comments were positive, compared to 26% of the spans. In fact, there were more total positive spans in negative comments ($n = 928$) than there were in positive comments ($n = 578$).

In work on review genres, it has been observed that negative evaluation tends to be preceded by some positive assessment, a type of ‘nice, but...’ structure (Taboada, Carretero & Hinnell, 2014). Commenters in our data, by contrast, do not shy away from starting negative: 45% of the comments contain only negative spans, and some comments open with negative spans (*This article is alarmist in the extreme*) and continue to negatively evaluate the article, with positive Attitude expressed only towards other views than those of the author of the article. Some other comments do avoid starting negative by opening with a suggestion (*We should do X...*) then follow up with negative appraisal (*...but currently we’re doing Y..., which we shouldn’t do*).

This overwhelmingly negative nature probably has to do with the characteristics of commenting on many social media sites and newspaper sites as well. Sites typically offer *reaction* options, such as the ‘like’ button on Facebook, the ‘heart’ on Twitter, a ‘Like’ option on the *Globe and Mail* website,⁸ or one of several other ways of sharing the content. We suspect that commenters who have a positive appraisal of the article simply use the ‘Like’ button. It is probably mostly commenters who disagree or are frustrated with the opinions in the article (or in other comments) that take the trouble to write in the comments section. This could help explain why so many of the comments are negative.

Appreciation and Judgement per comment were distributed roughly equally, matching our observation about the corpus in general. Within the average comment, it was about as common to use only Judgement, only Appreciation, or a balance of both, though there was a slight bias towards mostly using Appreciation with some Judgement rather than the reverse. This is reflected in the slightly greater frequency of Appreciation in the corpus overall.

6.3 Attitude and negation

The presence of negation undoubtedly affects the interpretation of evaluative expressions. The way in which this specifically takes place is a complex issue. Negation is intertwined with negativity, the former being a syntactic or lexical phenomenon, and the latter the semantic interpretation of negative words and negated statements. Potts (2011) shows a correlation between negation and negativity and characterizes negation as ‘persistently negative.’ See also Israel (2004) and Taboada, Trnavac & Goddard (2017). We examined, then, the relationship between syntactic negation and negative polarity in our annotations.

As part of the larger Corpus project, we annotated these same comments for negation, identifying (i) the negative keyword⁹ (*not*, *n’t*, and some lexical items such as *lack* or

8 The ‘Like’ option was available on the site of *The Globe and Mail* at the time we collected the data. The interface has changed since.

9 We used the term ‘keyword’ to refer to negative items that anchor the negation annotations, not in the technical sense of ‘keyword’ in corpus linguistics. The former meaning is common in studies of negation and negation annotation for computational purposes (Jiménez-Zafra *et al.*, 2021).

fail); (ii) the scope of the negation; and (iii) the focus of the negation (the word or phrase most directly affected by the negation). A full description of the annotation process and statistics on negation is provided in Kolhatkar *et al.* (2020). We isolated negation from the wider Engagement category within Appraisal, treating it as a syntactic phenomenon, although we are aware that it plays a role in the linguistic expression of Engagement. We addressed only its syntactic status because a full annotation of Engagement was beyond the scope of this project. The annotations for negation, however, are reliable and can contribute to our understanding of the expression of Attitude in the comments.

Once we had both Appraisal and negation annotations, we layered both sets of annotations, to extract Attitude spans that overlapped with the focus of grammatical negation. This was intended as a somewhat rough measure of finding the Attitude that is most directly affected by that negation. We choose focus instead of scope, because scope tends to be a much larger span.

We show an example of how the two annotations relate in (27). The first part contains the negation annotations. It is important to note here that, although the scope is the entire sentence after the negative keyword *lack*, the focus is only the word *awareness*. This is relevant because the Appraisal annotators (who performed the annotation independently of the negation annotators) annotated three different spans here. The only one that overlaps with the focus of negation is *lack of awareness*. The awareness is what is being negated, what is being presented as not in existence. Although the negation has scope over *intergenerational trauma* and *challenges*, those two Appraisal spans are in fact not negated, that is, their existence is not being put into question.

(27) Negation (a) and Appraisal (b) annotations for the same example

- a. ... it shows a [lack]_{Keyword} [of [awareness]_{Focus} of the impact of intergenerational trauma and the challenges it causes on the community.]_{Scope}
- b. ... it shows a [lack of awareness]_{Judg, neg} of the impact of [intergenerational trauma]_{App, neg} and the [challenges]_{App, neg} it causes on the community.

Figure 3 shows the distribution of Attitude polarities for spans overlapping with focus of negation. Relative to all spans in the corpus, those overlapping with focus of negation are more likely to be either negative or neutral. This is partially because neutral Attitude was only annotated when a commenter took a position that was explicitly neither positive nor negative; the usual way this happened was through the negation of negative Attitude.

Figure 4 shows the distribution of Attitude labels for spans overlapping with the focus of negation. Spans that overlap with focus of negation are more likely to be spans of Judgement, likely because Judgement in this corpus tends to be overwhelmingly negative (83% of Judgement spans are negative; see Table 3). One such example is in (28), where the long Judgement span that starts at *if the NDP had not joined the Conservatives* enacts

negative Judgement with the help of syntactic negation (*had not joined*). On the other hand, Affect is *less* likely to be in the focus of negation, even though Affect spans are also more frequently negative than positive (77% of the time). It seems that negative Affect is not always expressed through negation, but more often through negative words, whereas negative Judgement is more likely to be conveyed through negation. Spans that express negative Affect included *frightened*, *concerned*, *very sad*, or *sense of betrayal*, with only a few negating a positive (*don't appreciate* or *not upset at all*).

- (28) A great plan, but let's not forget that if the NDP had not joined the Conservatives in such a hurry to bring down the Liberal government in 2006 which had just established a national child care program (not promised, or planne4d, but established), Canada would have already had a national day care plan for EIGHT long years.

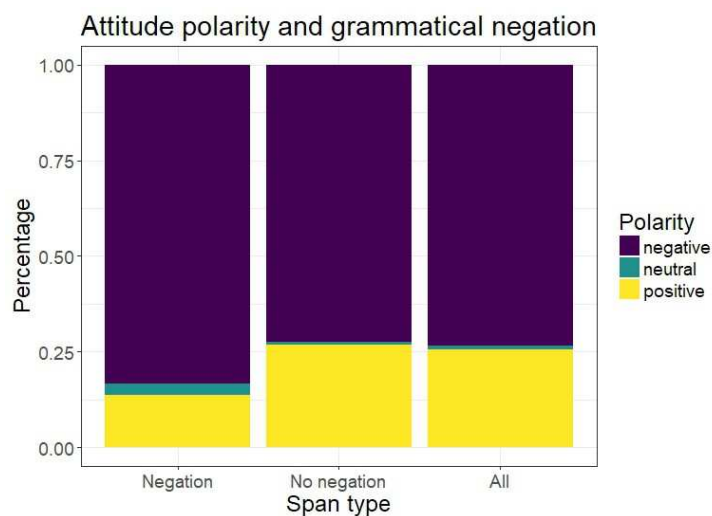


Figure 3: Distribution of polarities of spans of comments with negation, without negation, and undifferentiated

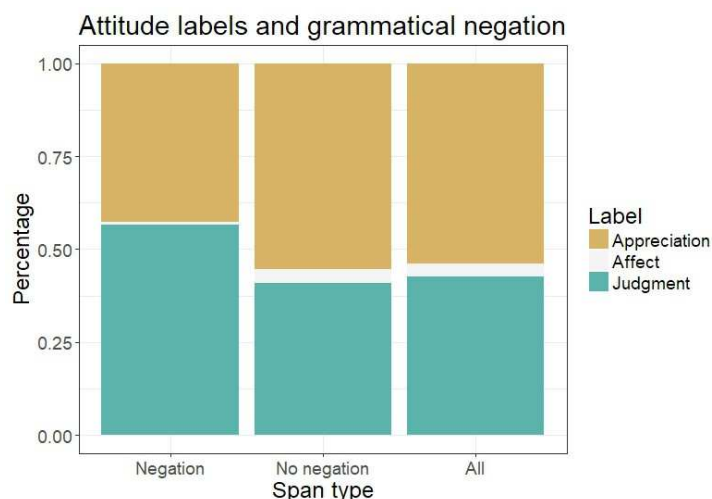


Figure 4: Distribution of labels of spans of comments with negation, without negation, and undifferentiated

6.4 Attitude and constructiveness

Another set of annotations that we carried out for this data involved assessing whether comments are ‘nice’ or ‘nasty’ in the context of online news. We defined those nice and nasty characteristics in terms of constructiveness and toxicity. Constructive comments are those that intend to create a civil dialogue through remarks that are relevant to the article and not intended to merely provoke an emotional response; they are typically targeted to specific points and supported by appropriate evidence. Toxic comments, on the other hand, are likely to offend or cause distress (Kolhatkar *et al.*, 2020). The constructiveness and toxicity annotations were slightly different from the Appraisal and negation annotations, as they were completed through crowdsourcing. We recruited workers on a crowdsourcing platform, provided definitions of the main concepts, and asked them to annotate individual comments (after having read the article each comment was responding to). Full details are provided in Kolhatkar *et al.* (2020). In Kolhatkar, Thain, Sorensen, Dixon & Taboada (to appear), we present a method to use such annotations to develop automatic methods to detect constructiveness and toxicity and to develop a system to moderate comments automatically, promoting those that are constructive and demoting the toxic ones.

We took the annotations, which were performed on a by-comment basis, and compared them to the Appraisal annotations within each comment. We were interested in whether constructive and/or toxic comments showed different Appraisal patterns.

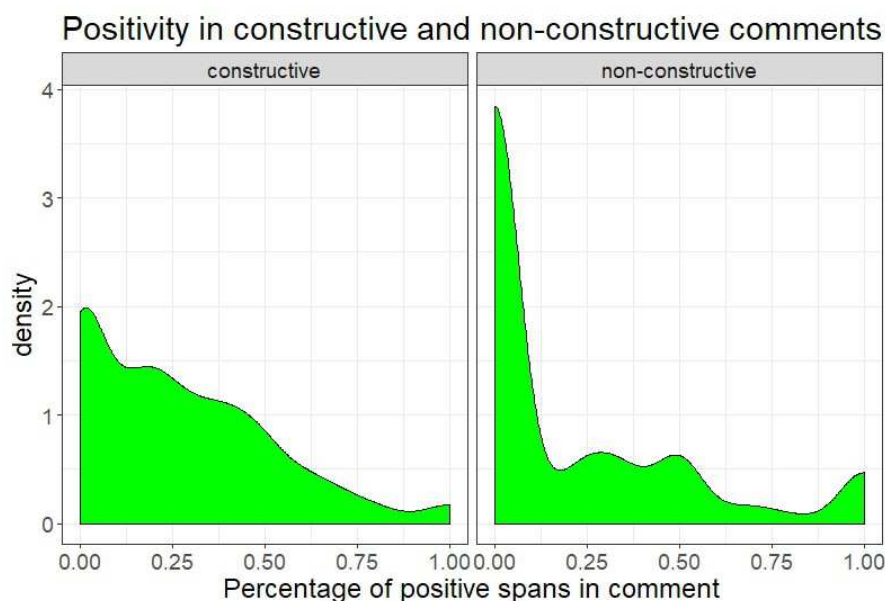


Figure 5: Density of constructiveness and positivity

In terms of constructiveness, we found that one indicator of a constructive comment is a mix of positive and negative spans. The proportion of constructive and non-constructive

comments that are mostly negative (those with more than 50% of their Appraisal spans annotated as negative) is nearly identical, rounding to 79%. Yet, as shown in Figure 5, a constructive comment is more likely to have some positive spans than a non-constructive comment. Therefore, mixing some positive Appraisal in with negative spans seems to be a mark of constructive comments. This likely gives some appearance of balance to these comments. There is no apparent corresponding difference in mostly positive comments, but that means little, as positive spans are so underrepresented in the corpus.

The presence of Affect is another marker of a constructive comment. Of all the comments, only 159 had some Affect spans in them. Within these, only 34 (21%) were annotated as non-constructive. Constructive comments with some Affect still use little Affect; they have a mean of merely 1.42 Affect spans (as opposed to their mean 7.21 Appreciation and 5.44 Judgement spans). Writers of such comments use Affect to describe both others' reactions and their own emotions regarding real or hypothetical events.

Graduation is also more common in constructive comments. Of the 398 comments with some Graduation spans, 289 (72%) appear in constructive comments.

6.5 Attitude and toxicity

The corpus was also annotated for toxicity through crowdsourcing, using a four-point scale: not toxic, mildly toxic, toxic, and very toxic (Kolhatkar *et al.*, 2020). We found that toxic comments were very rare, likely due to the fact that the *Globe's* platform includes moderation of comments, mostly automatic, but also through other users flagging comments for deletion. Of all the 1,043 comments, only 203 (19.46%) had some toxicity in them (as either mildly toxic, toxic, or very toxic).

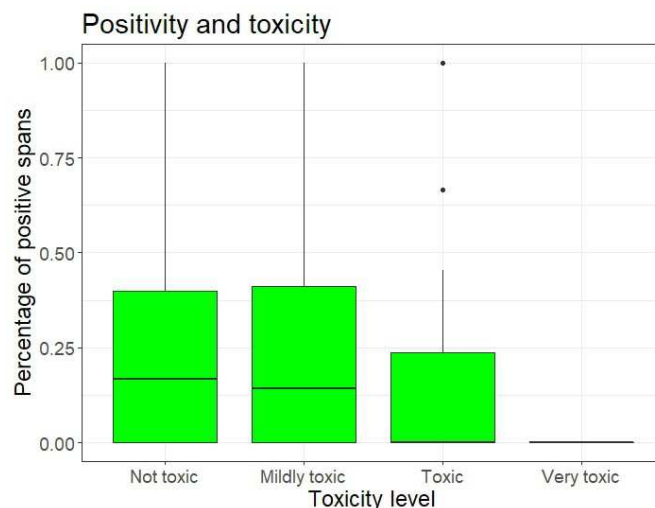


Figure 6: Comment toxicity and percentage of positive spans

Attitude label and polarity both seem to have a relationship with toxicity. Figure 6 shows that comments rated as toxic or very toxic have low rates of positive Attitude spans com-

pared to those labeled non-toxic or mildly toxic. However, as positive spans are generally uncommon in the corpus, span polarity is a weak indicator of toxicity.

Figure 7 shows the distribution of Appreciation and Judgement spans¹⁰ in comments at different levels of toxicity. In non-toxic comments, the median frequency of both Appreciation and Judgement is 50%, skewing slightly towards Appreciation. But in more toxic comments, Judgement spans are more frequent. A one-way analysis of variance (ANOVA) test confirms that the means for Appreciation and Judgement percentages are significantly different at different toxicity levels. For Appreciation, $F = 4.14$, adjusted $p < 0.01$; for Judgement, $F = 3.85$, $p < 0.01$.

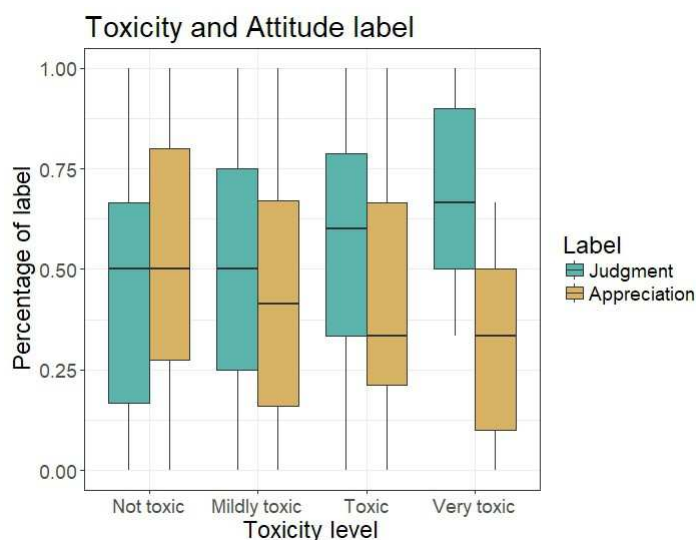


Figure 7: Comment toxicity and use of different Attitude labels

7. Discussion and conclusion

We have presented an analysis of Appraisal in a corpus of online news comments. The corpus contains 1,043 comments posted in relation to news stories on the website of the Canadian English-language newspaper *The Globe and Mail*. The Appraisal annotations were carefully carried out by two annotators, with one of them acting as curator for the entire corpus. Inter-annotator analyses show that the annotations (except for Graduation) are reliable and reproducible.

Our analyses include an overall characterization of Appraisal in this interesting genre, and the relationship of Appraisal to other phenomena that we have also annotated in the corpus.

First of all, and with regard to the characterization of Appraisal in this corpus, three main results emerge. The defining characteristic of the register of online news comments

¹⁰ Affect was excluded here due to its low frequency.

is their evaluative nature. The main function of comments is to evaluate the article, the ideas in the article and the people being discussed (politicians, public figures), in addition to evaluating the ideas of other commenters and other commenters themselves. This evaluation is predominantly negative: 73.5% of the spans in the corpus were negative. Finally, in terms of the different types of Attitude (Affect, Judgement, Appreciation), one surprising finding is that the frequency of Affect is quite low: Only 3.4% of the spans were labelled as Affect. We see this as surprising because we expected comments to express a strong emotion on the part of the commenter. Instead, strong emotional content is couched in terms of Judgement and Appreciation. In other words, rather than *I hate the candidate*, what we find is *The candidate is incompetent* or *The candidate's policies are bad*. This could be because commenters wish to convey some distance from their opinion.

Secondly, we studied the relationship between Appraisal and other aspects of the corpus: negation, constructiveness, and toxicity. With regard to Appraisal and negation, we found that Appraisal spans in the focus of negation are more likely to be either negative or neutral. We find negation in neutral spans, because that is precisely the type of cases where the polarity was difficult to settle. In terms of Attitude, spans in the focus of negation were more likely to be Judgement, likely because Judgement in general tends to be negative in our corpus, with 83% of Judgement spans being negative.

The comments overall were also annotated for constructiveness, that is, whether they contributed to the conversation and were meant to create a civil dialogue. Our analyses show that constructive comments tend to show a mix of positive and negative spans, rather than being exclusively either positive or negative. Constructive comments were more likely to express some Affect, although Affect is rare across the corpus. Predictably, non-constructive comments contained more negative spans.

A final set of annotations involved toxicity. Toxic comments in general were not frequent in our corpus, which consists of moderated comments. We did find that Judgement seems to be more prevalent in toxic comments than Appreciation, again highlighting the negative nature of Judgement in our corpus. Judgement is used when attacking individuals, whether the people mentioned in the article, the author of the article, or other commenters.

In sum, we find that the genre of online news comments seems to be more negative than other similar online genres, and that it seems to contain less Affect than we expected, less than other online genres such as Twitter discussions (Zappavigna, 2012). The analysis of Appraisal in our corpus presents a nuanced view of how online news comments deploy different types of Appraisal and how different Appraisal subtypes interact with negation, constructiveness, and toxicity.

Our results shed light into this new genre, which is beginning to be explored not just from a linguistic point of view, but also from the point of view of content moderation (Gillespie, 2018, 2020; Risch & Krestel, 2018; Seering, Wang, Yoon & Kaufman 2019).

The task of moderating comments, whether by automatic or manual means, involves making judgements on the language. When the language obscures subjectivity or negativity, as we have seen in our analyses, that task becomes more complex. In-depth, corpus-based analyses such as the ones presented here can help us better understand how evaluative language is expressed online and how to extract and analyze it for moderation tasks.

From a methodological point of view, we explore the application of a framework, Appraisal, which heavily relies on the researcher's intuitions and context knowledge, as a methodology to explore the discursive aspects of a corpus. The corpus, a subset of the much larger SFU Opinion and Comments Corpus, constitutes an instance of language in context, which has helped us discover the discourse properties of evaluative language in an online context. The results, naturally, apply only to this specific context (Canadian English, online news comments), but are likely indicative of the nature of comments in general, an underexplored area of research from the point of view of corpus-based discourse analysis.

Acknowledgements

This research was funded by the Social Sciences and Humanities Research Council of Canada (Insight Grant 435-2014-0171). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this project. We would like to thank members of the Discourse Processing Lab at Simon Fraser University for technical support and insightful discussions, and in particular Emilie Francis, Erin Jastrzebski, and Sarah Mulhall for their contribution to the Appraisal and negation annotations.

References

- Achugar, M. (2008). *What We Remember: The Construction of Memory in Military Discourse*. Amsterdam: John Benjamins. <https://doi.org/10.1075/dapsac.29>
- Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4), 555-596. <https://doi.org/10.1162/coli.07-034-R2>
- Baker, P. (2014). *Using Corpora to Analyze Gender*. London: Bloomsbury.
- Baker, P. (2020). Corpus-assisted discourse analysis. In C. Hart (Ed.), *Researching Discourse: A student guide* (pp. 124-142). New York: Routledge. <https://doi.org/10.4324/9780367815042-8>
PMCID:PMC7034358
- Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse Analysis and Media Attitudes: The representation of Islam in the British press*. Cambridge: Cambridge University Press.

- Bakhtin, M. (1981). Discourse in the novel (C. Emerson & M. Holquist, Trans.). In M. Holquist (Ed.), *The Dialogic Imagination: Four Essays by M. M. Bakhtin* (pp. 259-422). Austin: University of Texas Press.
- Barber, G. (2017, April 25). Readers help test Coral Project's commenting software, *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/ask-the-post/wp/2017/04/25/readers-help-test-coral-projects-commenting-software/?noredirect=on&utm_term=.42cec3135ac1
- Becker, A. (2009). Modality and Engagement in British and German political interviews. *Languages in Contrast*, 9(1), 5-22. <https://doi.org/10.1075/lic.9.1.02bec>
- Bednarek, M. (2006). *Evaluation in Media Discourse: Analysis of a newspaper corpus*. London: Continuum.
- Bednarek, M. (2009). Language patterns and ATTITUDE. *Functions of Language*, 16(2), 165-192. <https://doi.org/10.1075/fol.16.2.01bed>
- Bednarek, M. (2010). Evaluation in the news: A methodological framework for analyzing evaluative language in journalism. *Australian Journal of Communication*, 37(2), 15-50.
- Ben-Aaron, D. (2005). Given and news: evaluation in newspaper stories about national anniversaries. *Text*, 25(5), 691-718.
- Benamara, F., Taboada, M., & Mathieu, Y. (2017). Evaluative language beyond bags of words: Linguistic insights and computational applications. *Computational Linguistics*, 43(1), 201-264.
- Benesch, S. (2012). Words as weapons. *World Policy Journal*, 29(1), 7-12. <https://doi.org/10.1177/0740277512443794>
- Benítez Castro, M.-Á., & Hidalgo Tenorio, E. (2019). Rethinking Martin & White's affect taxonomy: A psychologically-inspired approach to the linguistic expression of emotion. In J. L. Mackenzie & L. Alba-Juez (Eds.), *Emotion in discourse* (pp. 301-331). Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.302.12ben>
- Berber Sardinha, T. (2018). Dimensions of variation across Internet registers. *International Journal of Corpus Linguistics*, 23(2), 125-157. <https://doi.org/10.1075/ijcl.15026.ber>
- Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511519871>

- Biber, D., & Egbert, J. (2016). Register variation on the searchable web. *Journal of English Linguistics*, 44(2), 95-137. <https://doi.org/10.1177/0075424216628955>
- Bilton, R. (2018). The Atlantic is killing its comments in favor of a new Letters section to showcase reader feedback. *NiemanLab*. <https://www.niemanlab.org/2018/02/the-atlantic-is-killing-its-comments-in-favor-of-a-new-letters-section-to-showcase-reader-feedback/>
- Clarke, I., & Grieve, J. (2019). Stylistic variation on the Donald Trump Twitter account: A linguistic analysis of tweets posted between 2009 and 2018. *PLoS ONE*, 14(9): e0222062. <https://doi.org/10.1371/journal.pone.0222062> PMID:31553740 PMCID:PMC6760825
- Coffin, C., & O'Halloran, K. (2006). The role of appraisal and corpora in detecting covert evaluation. *Functions of Language*, 13(1), 77-110. <https://doi.org/10.1075/fol.13.1.04cof>
- Dancygier, B., & Vandelanotte, L. (2017). Internet memes as multimodal constructions. *Cognitive Linguistics*, 28(3), 565. <https://doi.org/10.1515/cog-2017-0074>
- Davidson, T., Warmesley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language *Proceedings of the 11th International Conference on Web and Social Media* (pp. 512-515). Montréal.
- de Castilho, R. E., Mjrdicza-Maydt, E., Yimam, S. M., Hartmann, S., Gurevych, I., Frank, A., & Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)* (pp. 76-84). Osaka.
- de Jong, I., & Burgers, C. (2013). Do consumer critics write differently from professional critics? A genre analysis of online film reviews. *Discourse, Context and Media*, 31(2), 247-269. <https://doi.org/10.1016/j.dcm.2013.03.001>
- Diakopoulos, N. (2015). The editor's eye: Curation and comment relevance on the New York Times *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1153-1157). Vancouver. <https://doi.org/10.1145/2675133.2675160>
- Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., . . . Danforth, C. M. (2015). Human language reveals a positivity bias. *Proceedings of the National Academy of Sciences*, 112(8), 2389-2394.

- <https://doi.org/10.1073/pnas.1411678112>
PMid:25675475 PMCID:PMC4345622
- Dotti, F. C. (2013). Overcoming problems in automated appraisal recognition: The attitude system in inscribed appraisal. *Procedia - Social and Behavioral Sciences*, 95, 442-446. <https://doi.org/10.1016/j.sbspro.2013.10.667>
- Eggins, S., & Slade, D. (1997). *Analysing Casual Conversation*. London: Cassell.
- Ehret, K., & Taboada, M. (2020). Are online news comments like face-to-face conversation? A multi-dimensional analysis of an emerging register. *Register Studies*, 2(1), 1-36. <https://doi.org/10.1075/rs.19012.ehr>
- Ehret, K., & Taboada, M. (to appear). Characterising online news comments: A multi-dimensional cruise through online registers. *Frontiers in Artificial Intelligence*.
- Farina, M. (2018). *Facebook and Conversation Analysis: The structure and organization of comment threads*. London: Bloomsbury Publishing.
- Flowerdew, L. (2013). Corpus-based discourse analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 174-187). New York: Routledge.
- Fuoli, M. (2012). Assessing social responsibility: A quantitative analysis of Appraisal in BP's and IKEA's social reports. *Discourse & Communication*, 6(1), 55-81. <https://doi.org/10.1177/1750481311427788>
- Fuoli, M. (2018). A step-wise method for annotating APPRAISAL. *Functions of Language*, 25(2), 229-258. <https://doi.org/10.1075/fol.15016.fuo>
- Fuoli, M., & Hommerberg, C. (2015). Optimising transparency, reliability and replicability: Annotation principles and inter-coder agreement in the quantification of evaluative expressions. *Corpora*, 10(3), 315-349. <https://doi.org/10.3366/cor.2015.0080>
- Gagnon, L. (2016). Why Belgium is ground zero for jihadi terrorism. *The Globe and Mail*. <https://www.theglobeandmail.com/opinion/why-belgium-is-ground-zero-for-jihadi-terrorism/article29425530/>
- Gardiner, B., Mansfield, M., Anderson, I., Holder, J., Louter, D., & Ulmanu, M. (2016). The dark side of Guardian comments. *The Guardian*, April 12, 2016. <https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments>

- Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven: Yale University Press.
- Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*, 7(2), 2053951720943234. <https://doi.org/10.1177/2053951720943234>
- Giltrow, J., & Stein, D. (2009). *Genres in the Internet: Issues in the theory of genre*. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.188>
- Halliday, M. A. K. (1985). *An Introduction to Functional Grammar* (1st ed.). London: Arnold.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar* (3rd ed.). London: Arnold. <https://doi.org/10.4324/9780203783771>
- Hardaker, C. (2015). 'I refuse to respond to this obvious troll': An overview of responses to (perceived) trolling. *Corpora*, 10(2), 201-229. <https://doi.org/10.3366/cor.2015.0074>
- Hardaker, C., & McGlashan, M. (2016). "Real men don't hate women": Twitter rape threats and group identity. *Journal of Pragmatics*, 91, 80-93. <https://doi.org/10.1016/j.pragma.2015.11.005>
- Herring, S., Scheidt, L. A., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs *Proceedings of the 37th Annual Hawaii International Conference on System Sciences* (pp. 11 pp.). Hawaii. <https://doi.org/10.1109/HICSS.2004.1265271>
- Hommerberg, C., & Don, A. (2015). Appraisal and the language of wine appreciation: A critical discussion of the potential of the Appraisal framework as a tool to analyse specialised genres. *Functions of Language*, 22(2), 161-191. <https://doi.org/10.1075/fol.22.2.01hom>
- Horvath, B. M., & Eggins, S. (1995). Opinion texts in conversation. In P. H. Fries & M. Gregory (Eds.), *Discourse in Society: Systemic Functional Perspectives* (pp. 29-45). Norwood, NJ: Ablex.
- Israel, M. (2004). The pragmatics of polarity. In L. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 701-723). Malden, MA: Blackwell. <https://doi.org/10.1002/9780470756959.ch31>
- Jane, E. A. (2016). *Misogyny Online: A short (and brutish) history*. Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781473916029>

- Jiménez-Zafra, S. M., Cruz-Díaz, N. P., Taboada, M., & Martín-Valdivia, M. T. (2021). Negation detection for sentiment analysis: A case study in Spanish. *Natural Language Engineering*, 27(2), 225-248. <https://doi.org/10.1017/S1351324920000376>
- Kiesling, S. F., Pavalanathan, U., Fitzpatrick, J., Han, X., & Eisenstein, J. (2018). Interactional stancetaking in online forums. *Computational Linguistics*, 44(4), 683-718. https://doi.org/10.1162/coli_a_00334
- Kolhatkar, V., Thain, N., Sorensen, J., Dixon, L., & Taboada, M. (to appear). Classifying constructive comments. *First Monday*.
- Kolhatkar, V., Wu, H., Cavasso, L., Francis, E., Shukla, K., & Taboada, M. (2018). *The SFU Opinion and Comments Corpus: A corpus for the analysis of online news comments*. Retrieved from: <https://github.com/sfu-discourse-lab/SOCC>
- Kolhatkar, V., Wu, H., Cavasso, L., Francis, E., Shukla, K., & Taboada, M. (2020). The SFU Opinion and Comments Corpus: A corpus for the analysis of online news comments. *Corpus Pragmatics*, 4(2), 155-190. <https://doi.org/10.1007/s41701-019-00065-w>
PMid:32685909 PMCID:PMC7357677
- Kwok, I., & Wang, Y. (2013). Locate the hate: Detecting tweets against Blacks *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 27, pp. 1621-1622). Bellevue.
- Lam, S. L., & Crosthwaite, P. (2018). Appraisal resources in L1 and L2 argumentative essays: A contrastive learner corpus-informed study of evaluative stance. *Journal of Corpora and Discourse Studies*, 1(1), 8-35.
- Llansó, E. J. (2020). No amount of “AI” in content moderation will solve filtering’s prior-restraint problem. *Big Data & Society*, 7(1), January. <https://doi.org/10.1177/2053951720920686>
- Loosen, W., Häring, M., Kurtanović, Z., Merten, L., Reimer, J., van Roessel, L., & Maalej, W. (2018). Making sense of user comments: Identifying journalists’ requirements for a comment analysis framework. *SCM Studies in Communication and Media*, 6(4), 333-364. <https://doi.org/10.5771/2192-4007-2017-4-333>
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 157-176). Amsterdam: Benjamins.
- Macken-Horarik, M. (2003). APPRAISAL and the special instructiveness of narrative. *Text*, 23(2), 285-312. <https://doi.org/10.1515/text.2003.012>

- Macken-Horarik, M., & Isaac, A. (2014). Appraising appraisal. In G. Thompson & L. Alba-Juez (Eds.), *Evaluation in context* (pp. 67-92). Amsterdam: John Benjamins.
<https://doi.org/10.1075/pbns.242.04mac>
- Manosevitch, I., & Tenenboim, O. (2017). The multifaceted role of user-generated content in news websites. *Digital Journalism*, 5(6), 731-752.
<https://doi.org/10.1080/21670811.2016.1189840>
- Marcoccia, M. (2004). On-line polylogues: conversation structure and participation framework in internet newsgroups. *Journal of Pragmatics*, 36(1), 115-145.
[https://doi.org/10.1016/S0378-2166\(03\)00038-9](https://doi.org/10.1016/S0378-2166(03)00038-9)
- Martin, J. R. (1984). Language, register and genre. In F. Christie (Ed.), *Children Writing: Reader* (pp. 21-30). Geelong, Victoria: Deakin University Press.
- Martin, J. R. (2000). Beyond exchange: Appraisal systems in English. In S. Hunston & G. Thompson (Eds.), *Evaluation in Text: Authorial Distance and the Construction of Discourse* (pp. 142-175). Oxford: Oxford University Press.
- Martin, J. R. (2017). The discourse semantics of attitudinal relations: Continuing the study of lexis. *Russian Journal of Linguistics*, 21(1), 22-47.
<https://doi.org/10.22363/2312-9182-2017-21-1-22-47>
- Martin, J. R., & Rose, D. (2008). *Genre Relations: Mapping Culture*. London: Equinox.
- Martin, J. R., & White, P. R. R. (2005). *The Language of Evaluation*. New York: Palgrave.
<https://doi.org/10.1057/9780230511910>
- McGuire, J. (2015). Uncivil dialogue: Commenting and stories about indigenous people. *CBC News*, November 30, 2015.
<http://www.cbc.ca/newsblogs/community/editorsblog/2015/11/uncivil-dialogue-commenting-and-stories-about-indigenous-people.html>
- McKinney, W. (2010). Data structures for statistical computing in python *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51-56). Austin, TX.
<https://doi.org/10.25080/Majora-92bf1922-00a>
- Meyer, H. K., & Carey, M. C. (2014). In moderation: Examining how journalists' attitudes toward online comments affect the creation of community. *Journalism Practice*, 8(2), 213-228. <https://doi.org/10.1080/17512786.2013.859838>
- Meyer, H. K., & Speakman, B. (2016). Quieting the commenters: The spiral of silence's persistent effect on online news forums. *ISOJ Journal*, 6(1), April 14.

- Mishra, P., Del Tredici, M., Yannakoudakis, H., & Shutova, E. (2019). *Abusive language detection with graph convolutional networks*. Paper presented at the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN.
- Nakamura, L. (2015). The unwanted labour of social media: Women of colour call out culture as venture community management. *New Formations*, 86(86), 106-112. <https://doi.org/10.3898/NEWF.86.06.2015>
- Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive language detection in online user content *Proceedings of the 25th International Conference on World Wide Web* (pp. 145-153). Montréal. <https://doi.org/10.1145/2872427.2883062>
- O'Reilly, T. (2005). What is Web 2.0: Design patterns and business models for the next generation of software. *O'Reilly Network*. Retrieved from O'Reilly Network website: www.oreilly.com/pub/a/web2/archive/what-is-web-20.html
- Partington, A. (2014). Evaluative prosody. In K. Aijmer & C. Rühleman (Eds.), *Corpus pragmatics: A handbook* (pp. 279-303). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139057493.015>
- Partington, A., Morley, J., & Haarman, L. (Eds.). (2004). *Corpora and Discourse*. Bern: Peter Lang.
- Potts, C. (2011). On the negativity of negation *Proceedings of SALT 20: Semantics and Linguistic Theory* (pp. 636-659). Vancouver. <https://doi.org/10.3765/salt.v20i0.2565>
- R Core Team (2018). R: A Language and Environment for Statistical Computing: R Foundation for Statistical Computing, <https://www.R-project.org>.
- Read, J., & Carroll, J. (2012). Annotating expressions of Appraisal in English. *Language Resources and Evaluation*, 46, 421-447.
- Reagle, J. M. (2015). *Reading the comments: Likers, haters, and manipulators at the bottom of the Web*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/10116.001.0001>
- Reich, Z. (2011). User comments: The transformation of participatory space. In J. B. Singer, A. Hermida, D. Domingo, A. Heinonen, S. Paulussen, T. Quandt, Z. Reich & M. Vujnovic (Eds.), *Participatory journalism: Guarding open gates at online newspapers* (pp. 96-117). Hoboken, NJ: Wiley-Blackwell. <https://doi.org/10.1002/9781444340747.ch6>
PMid:21543057

- Risch, J., & Krestel, R. (2018). *Delete or not delete? Semi-automatic comment moderation for the newsroom*. Paper presented at the Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), Santa Fe.
- Risch, J., Repke, T., Kohlmeyer, L., & Krestel, R. (2021). ComEx: Comment exploration on online news platforms *Joint Proceedings of the ACM IUI 2021 Workshops co-located with the 26th ACM Conference on Intelligent User Interfaces (IUI)* (pp. 1-7). College Station, TX (online).
- Roberts, S. T. (2019). *Behind the Screen: Content moderation in the shadows of social media*. New Haven: Yale University Press. <https://doi.org/10.2307/j.ctvhrcz0v>
- Seering, J., Wang, T., Yoon, J., & Kaufman, G. (2019). Moderator engagement and community development in the age of algorithms. *New media & society*, 21(7), 1417-1443. <https://doi.org/10.1177/1461444818821316>
- Skalicky, S. (2013). Was this analysis helpful? A genre analysis of the Amazon.com discourse community and its “most helpful” product reviews. *Discourse, Context and Media*, 84-93. <https://doi.org/10.1016/j.dcm.2013.04.001>
- Sobieraj, S. (2020). *Credible Threat: Attacks against women online and the future of democracy*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780190089283.001.0001>
- Stewart, D. (2010). *Semantic Prosody: A Critical Evaluation*. New York: Routledge. <https://doi.org/10.4324/9780203870075>
- Stroud, N. J., Murray, C., & Kim, Y. (2020). News comments: What happens when they're gone or when newsrooms switch platforms. Austin, TX: Center for Media Engagement.
- Taboada, M. (2011). Stages in an online review genre. *Text and Talk*, 31(2), 247-269. <https://doi.org/10.1075/lic.14.1.07tab>
- Taboada, M., Carretero, M., & Hinnell, J. (2014). Loving and hating the movies in English, German and Spanish. *Languages in Contrast*, 14(1), 127-161.
- Taboada, M., & Grieve, J. (2004). Analyzing appraisal automatically. In Y. Qu, J. G. Shanahan & J. Wiebe (Eds.), *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07)* (pp. 158-161). Stanford University, CA: AAAI Press.

- Taboada, M., Trnavac, R., & Goddard, C. (2017). On being negative. *Corpus Pragmatics*, 1(1), 57-76. <https://doi.org/10.1007/s41701-017-0006-y>
- Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020). The dynamics of political incivility on Twitter. *Sage Open*, 10(2), 1-15. <https://doi.org/10.1177/2158244020919447>
- Thompson, G. (2014). Affect and emotion, target-value mismatches, and Russian dolls: Refining the Appraisal model. In G. Thompson & L. Alba-Juez (Eds.), *Evaluation in context* (pp. 47-66). Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.242.03tho>
- Turpel-Lafond, M. E. (2014). Enough is enough: Time to address epidemic of violence against native women. *The Globe and Mail*. <https://www.theglobeandmail.com/opinion/put-native-women-on-the-agenda/article20144737/>
- Vásquez, C. (2014). *The Discourse of Online Consumer Reviews*. London: Bloomsbury.
- Veletsianos, G., Houlden, S., Hodson, J., & Gosse, C. (2018). Women scholars' experiences with online harassment and abuse: Self-protection, resistance, acceptance, and self-blame. *New media & society*, 20(12), 4689-4708. <https://doi.org/10.1177/1461444818781324>
- Warner, W., & Hirschberg, J. (2012). Detecting hate speech on the World Wide Web *Proceedings of the Second Workshop on Language in Social Media* (pp. 19-26). Montréal, Canada.
- Waseem, Z., & Hovy, D. (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter *Proceedings of NAACL-HLT* (pp. 88-93). San Diego, CA. <https://doi.org/10.18653/v1/N16-2013>
- White, P. R. R. (1998). *Telling Media Tales: The News Story as Rhetoric*. (Ph.D. dissertation), University of Sydney, Sydney.
- White, P. R. R. (2002). Appraisal: The language of evaluation and stance. In J.-O. Östman & J. Verschueren (Eds.), *Handbook of Pragmatics* (pp. 1-27). Amsterdam: John Benjamins.
- White, P. R. R. (2003). Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text*, 23(2), 259-284. <https://doi.org/10.1515/text.2003.011>

- White, P. R. R. (2016). Evaluative contents in verbal communication. In A. Rocci & L. De Saussure (Eds.), *Verbal Communication* (pp. 77-96). Berlin: Walter de Gruyter. <https://doi.org/10.1515/9783110255478-006>
- Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. New York: Springer. <https://doi.org/10.1007/978-0-387-98141-3>
- Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the machine: Anti-Black and Anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology*, 60(1), 93-117. <https://doi.org/10.1093/bjc/azz064>
- Wolfgang, J. D. (2018). Cleaning up the “Fetid Swamp”. *Digital Journalism*, 6(1), 21-40. <https://doi.org/10.1080/21670811.2017.1343090>
- Wright, L., Ruths, D., Dillon, K. P., Saleem, H. M., & Benesch, S. (2017). Vectors for counterspeech on Twitter *Proceedings of the First Workshop on Abusive Language Online* (pp. 57-62). Vancouver, BC. <https://doi.org/10.18653/v1/W17-3009>
- Wulczyn, E., Thain, N., & Dixon, L. (2017). Ex machina: Personal attacks seen at scale *Proceedings of the 26th International Conference on the World Wide Web* (pp. 1391-1399). Perth, Australia. <https://doi.org/10.1145/3038912.3052591>
- Yates, J., & Orlikowski, W. (1992). Genres of organisational communication: An approach to studying communication and media. *The Academy of Management Review*, 17(2), 299-326. <https://doi.org/10.5465/amr.1992.4279545>
- Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New media & society*, 13(5), 788-806. <https://doi.org/10.1177/1461444810385097>
- Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web*. London: Bloomsbury.